Evaluating Teaching

DISCUSSION GUIDE

P.A. Cranton
and
G.L. Geis

McGill
University

"EVALUATING TEACHING"

DISCUSSION GUIDE

P. A. Cranton

and

G. L. Geis

Centre for Teaching and Learning Services
McGill University
Montreal, Quebec
Canada
1982

Acknowledgements

This set of videotapes and guide is the result of long and close cooperation of many people. We wish to thank:

- Our McGill colleagues Dale Dauphinee, Irwin Gopnik and Tom Velk
- Concordia Professors Ron Smith and Phil Abrami. The latter was responsible for bringing to McGill the five evaluation specialists seen in Tapes 2 and 3.
- Our visiting experts: Lawrence Aleamoni, John Centra, Kenneth Doyle, Wilbert McKeachie and Robert Menges.
- Myrna Gopnik, Director of University Teaching and Learning, for her continued support.
- The outstanding staff of McGill's Instructional Communication Centre. These include our patient and talented producer-directors, Mark Sandiford and Mario di Paolo and their crew, and graphic artists Wim Van Eyk, Max Stiebel, Michael Coté.
- Our uncomplaining and devoted typists, Marie LaRicca and Anne Sage.
- And our many other colleagues and students who contributed suggestions, advice and criticism.

Table of Contents

Introduction	4
Videotapes	4
Guide	4
Using the Package	5
Specific Suggestions	5
Tape No. 1: Basic Issues and Concerns	7
Tape No. 2: The Student Response Questionnaire	15
Tape No. 3: Other Sources of Information	35
Tape No. 4: Putting It All Together	44
Selected General Bibliography	48

Introduction

The "package" consists of four videotapes and this Guide.

The purpose of this set of materials is to provide information to faculty groups on college and university campuses, to promote informed discussion about the evaluation of teaching, and to offer a model for developing or revising an evaluation plan.

1. The Videotapes

Tape No. 1, "Basic Issues and Concerns," is a discussion with faculty members of several basic issues which should be considered at some length by any group planning to set up or revise an evaluation system.

Tape No. 2, "The Student Response Questionnaire," offers a close look at one particular method used to gather information for evaluating teaching. Five evaluation specialists are interviewed; they comment on a wide variety of matters, from technical to philosophical concerns.

Tape No. 3 is entitled "Other Sources of Information." It continues the discussion with the experts who now comment on potential sources of information other than student questionnaires.

Tape No. 4, "Putting It All Together." In a simulated interview with a faculty member, this tape outlines a method for devising and implementing a systematic, comprehensive evaluation plan.

2. The Guide

This booklet is designed to help structure the use of the tape programs as generators of discussion and to provide resources for further study.

The section of the guide related to each videotape consists of a set of discussion questions and comments focused on key concepts and controversial points suggested by the videotape and includes selected references to the

literature. At the end of the guide is a short annotated bibliography representing a basic library on evaluating teaching.

3. Using the package

We feel strongly that if the evaluation effort is to be a serious one and if evaluation is to be seen as an integral, continuing part of the teaching function, the faculty as a whole must be involved from the outset. Tough questions and strong disagreements should come out in the open early in the process. Otherwise they remain as time bombs which will periodically go off, damaging or even destroying the evaluation effort.

There is no substitute for generating a series of open discussions involving all of the persons to be evaluated. These films and guide are designed to promote and structure such discussions. On many campuses there is a faculty development office or a faculty committee which might take the lead in presenting these films to groups.

4. Specific suggestions

(a) We suggest, especially in larger institutions, that these discussions be carried out at the level of faculties or departments to encourage participation of all the members of the academic community and to examine concerns specific to each group. Since the subject is seen as threatening to some, an informal atmosphere is to be recommended at least until specific tasks are set.

(b) The purposes of the meetings should be made explicit. If purposes change over time (say, from a general overview of teaching evaluation to the specific task of developing a departmental evaluation plan), that shift should be made overt and agreed to.

(c) We suggest a distribution of the copies of the appropriate section of this Guide before each meeting. (Users are permitted to photocopy these

materials).

(d) At least an hour should be devoted to each presentation. The programs are of different lengths ranging from 10 - 45 minutes. With an adequate introduction and discussion after the showing, an hour will easily be consumed.

(e) Although the tapes were planned as a sequence, each can be shown independently of the others; however, tapes No. 2 and No. 3 should always be used as a pair.

(f) To set the stage and open up discussion, Tape No. 1 might be presented. However, it may be that the group is not yet ready to tackle such concerns. Tapes No. 2 and No. 3, which deal specifically with gathering data for evaluation, should then be used as openers. After sessions based on those tapes, the more personal matters, underlying assumptions, and concerns may start to surface. Then Tape No. 1 would be timely for viewing at the next session.

(g) Tape No. 4 simulates, in a condensed time span, the series of steps a group might go through in developing an evaluation plan. This unit might be a good overview for faculty members before work on a plan actually begins so that everyone sees the sequence of events and understands the purpose and place of each step. In some cases it might be the first tape shown and discussed.

(h) Much of the discussion is likely to involve differing opinions on matters which in fact have been studied by researchers. The Guide offers a window into that literature. We suggest that the references in the Guide be obtained and made available before, during, and after each session.

Tape No. 1: Basic Issues and Concerns

Running time: 32 minutes.

In this roundtable discussion of evaluating teaching several university professors examine a number of major philosophical and practical concerns frequently expressed by faculty members. Centre staff members act as moderator and summarizers.

We suggest that this is a good introductory tape which can lead off discussions in your group. However, as we stated earlier, if the group is focussed upon particular details of evaluation (e. g. student questionnaires) it might be better to start with Tapes 2 and 3 and return to this tape as philosophical concerns emerge.

1. Can teaching be evaluated?

Frequently professors state that teaching cannot be evaluated. But pursuing that remark often reveals that few mean it literally. Only a small number of professors seriously contend that teaching is ineffable -- a mysterious art beyond observation and judgement. Other arts, painting for example, are evaluated in terms of both process and product by critics, collectors, and artists themselves.

More often the professor means that teaching should not be evaluated. Or he or she is expressing a concern about <u>why</u>, and <u>how</u> it should be evaluated, <u>what</u> should be evaluated, and by <u>whom</u>.

2. What should be the purpose of the evaluation?

Professors report that they regularly engage in self evaluation for the purposes of monitoring and improving their teaching. Students periodically issue course guides which often contain questionnaire data. However it is the third use of evaluation, for administrative decision making (e. g.

promotion and tenure decisions) that is usually the most troublesome.

Professors are often concerned that information about their teaching will be disseminated widely and will be misused; confidentiality is a serious problem.

Consequently, it is of primary importance that the purpose for which information is being collected is made clear at the outset and that adequate safeguards are set up to insure that the data will not be used for other purposes.

Even if these safeguards are in place, professors understandably raise the other questions previously mentioned.

3. How will evaluation information be gathered and how will it be used?

"Fairness" is a frequently raised concern. Is the process of gathering the data ethically acceptable and are the techniques technically sound. Is the evaluation comprehensive; does it include many sources of data and cover a sizeable time span? Does it deal somehow with biases? (Mulitiple sources of data can help offset biases somewhat). Does it provide sufficient opportunity for the professor to make his or her own case and to comment upon incomplete or seemingly biased information?

Even if information is gathered in acceptable ways there is concern about how it could be used. What are the criteria to be applied? How are data interpreted? What weight will the evaluation of teaching carry? In short, what are the rules and guidelines for decision making? As the speakers note in this tape, some fears about evaluation seem to arise from the mysteries surrounding decision-making. An early item on the discussion agenda should be explication of the decision process in which evaluation

information will be used.

4. Who uses evaluation information?

This point was raised earlier. Those who work in faculty development activities on campus are especially sensitive to the need for confidentiality when obtaining information for purposes of improvement. The audience to whom the information will be made available not only influences the degree of resistance by professors but also may affect the inputs of those who contribute to that information. For example, there is some evidence that students respond to a course questionnaire differently if it is to be used for teaching improvement than if it is to be used in promotion and tenure decisions.

In addition the professor is likely to be concerned about the abilities, biases and sophistication of those who will interpret the data. The discussions on this tape and those that follow repeatedly caution about responding to numerical data naively as if numbers themselves had magical qualities. "Fairness" requires that the interpreters of information be technically competent. Statisticians and others on a campus might well be consulted.

5. What should be evaluated?

Inevitably discussions focus almost at once on classroom teaching performance. But "teaching" is not properly defined solely in terms of classroom performance. Professors carry out a number of other activities such as evaluation of student learning and preparation of written materials. Furthermore they counsel students, meet with individual students as they prepare papers and conduct experiments, serve on curriculum committees and do many other things outside of the classroom that ought to be considered relevant to the teaching function. A comprehensive evaluation plan should

include gathering information on all such activities.

Some people would contend that what ought to be evaluted is learning — the outcome of the teaching enterprise (see, e. g. Rose, 1976 and McNeil & Popham, 1973). Several serious concerns arise: How is learning to be measured? And, when? To what degree can the teacher be seen as a factor in student achievement or change? (After all, students arrive with varying amounts of interest, motivation, prerequisite skills and knowledge. On the other hand, if there is little evidence of learning having taken place what was the point of teaching?)

One compromise position, borrowed from other professions, suggests judging the professional behaviors of the professor rather than outcomes. This is analogous to asking if the physician carried out at least the agreed upon minimum of professional activities in a particular case regardless of the final outcome in terms of the patient's health. While this is an interesting approach it raises questions about what these minimal teaching activities are.

6. What should be the contribution of students to the evaluation of teaching?

One need not conceive of evaluation solely in terms of student input. Indeed, some would argue for its total exclusion. In Tape 2 there is general agreement among the evaluation specialists that some sort of information from students is required for a comprehensive evaluation.

Students may provide input even at the early stages of planning (e. g., student members of an evaluation committee). They can offer valuable advice if a student reponse questionnaire is being developed.

Most often the challenge is to the nature of the data produced by such questionnaires. What kind of valid information can students supply? About what elements of teaching should they comment? How easily are students

biased by grades, teacher popularity, time of course presentations, etc? Won't students be more positive about courses they choose to take? Do students really know at the end of a course how much they have learned? How relevant it was? How good it was?

Most of these questions have been the subject of extensive research. Tape 2 and the accompanying section of this guide address many of them. The main point to be made here is that major validity problems have been studied and results of those studies are available in the literature on evaluation. We suggest that faculty members become informed about this area as they would any special area in their own discipline. This is not to suggest that there are simple answers readily available but rather to urge faculty members to discuss these complex questions in the light of what has been found out by researchers.

Much of the literature assumes that student reactions are critical to comprehensive evaluation. Here are two articles which strongly question the use of student opinion.

Kerlinger, F. N. Student evaluation of university professors. School and Society. October, 1971, 99, 353-356.

Renner, Richard, R. Comparing professors: how student ratings contribute to the decline in quality of higher education. Phi Delta Kappan. 1981, 83, 128-130.

7. How do you compare professors?

Despite the fact that comparisons may be odious, it seems natural to compare one professor with another. Sometimes comparisons are made among the faculty members of one department or unit, sometimes all members of a campus are compared. Occassionally, though rarely, comparisons are made with professors of similar courses at other institutions.

One can question any comparisons at all. Courses, teaching styles, students all differ. There are serious questions (some of which are raised in later tapes) about the validity of comparisons, given the wide array of possibly biasing factors. However, evaluation does imply the comparison of a specific set of data with some standards. "How good?" can only be answered in terms of "Relative to what?"

Alternatives to comparing professors have been suggested: an ideal "target" profile or a "gain score."

A template representing the ideal might be placed over the actual profile of a professor and the amount of difference noted (so-called Discrepancy Evaluation). But, who is to construct the ideal? Some suggest developing the model from a set of good teaching characteristics. That assumes there is an agreed upon list of such characteristics and there is some agreement in the literature about these "traits" (see, e.g. Cohen, et al, 1973). Others have suggested that the professor generate a good (or perfect) teacher profile and note how closely he or she approximates this self-generated goal. That path has other obvious pitfalls.

A different tack would emphasize growth and improvement. Evaluation would involve looking at positive changes (of what? and in what direction?) One problem with this approach is a so-called ceiling effect: the excellent teacher has little room in which to show improvement. And furthermore a great deal of growth may be exhibited by a teacher who moves from "terrible" to "very bad."

There is simply no escape from the difficult task of defining criteria or attributes which all or most teachers agree are important. This topic is returned to for discussion on succeeding tapes.

A quick answer to the problem can do real injustice to some things we

do know about teaching. Professors have very different models of teaching and its related activities and its possibilities (see, for example, Eble 1972). These models range all the way from a presenter of information -- a living book, if you will -- to an agent of personal student growth. Obviously what each of these professors would consider important and how much he or she would weight each attribute would differ dramatically. To impose a single model of teaching (as a single set of evaluation criteria implies) may not only provoke strong faculty reaction but also in the long run reduce the variety and quality of teaching.

This is not an issue to be resolved in a single stroke; it requires repeated visiting. The discussions about it are likely to be revealing and productive.

Some General References

There is a large literature on evaluating teaching. Here are a few items which we have found to be especially useful.

Centra, J. A. Determining faculty effectiveness. San Francisco: Jossey-Bass, 1980.

Cohen, A.M., Trent, J.W. and Rose, C. "Evaluation of Teaching" pps. 1041-1052 in Second Handbook of Research on Teaching (R.M.W. Travers, Ed.). Chicago: Rand McNally, 1973.

Eble, K. E. Professor as Teacher. San Francisco: Jossey Bass, 1972.

Eble, K. E. (Ed.). Improving teaching styles. San Francisco: Jossey-Bass, 1980.

Knapper, C., et al. (eds.). If Teaching is important . . . . . the evaluation of instruction in higher education. Toronto: Clarke, Irwin, 1979.

McNeil, J. D. & Popham W. S. The assessment of teacher competence. In R. M. W. Travers (Ed.), Second Handbook of Research on Teaching.

Chicago: Rand McNally, 1973.

Murray, H. G. Evaluating university teaching. Toronto: Ontario. Confederation of University Faculty Associations, 1980.

"Procedures for Evaluating instruction in a University setting." Report prepared by the Committee for the Improvement of Teaching and Learning, University of Alberta, Feb. 8, 1979, 11 pps.

Rose, C. Stalking the perfect teacher, Chronicle of Higher Education, 1981, 63, (2), 128-130.

Tape No.2: The Student Response Questionnaire

Running Time: 45 minutes with break at 25 minutes.

This tape addresses many of the complex issues involved in the development, use, and interpretation of student questionnaires in the evaluation of instruction. The questionnaire is the most commonly used instrument for collecting evaluation information from students and is also a technique which has been well-researched. Consequently a discussion of the questionnaire is likely to touch on a wide variety of issues.

This videotape is the longest in the series; a break point has been indicated on the tape. You may wish to view the tape in two sessions, providing time for discussion after each of the two parts.

This section of the discussion guide presents sixteen commonly raised issues related to the use of student questionnaires. For each, there is a brief description of the issue, questions which you may want to discuss as a group, and references which can be consulted for further information.

1. Students as evaluators of instruction

To many people it seems obvious that some kind of information from students ought to be included in any evaluation of teaching. After all, students are there in the classroom; they are assigned the readings; they carry out laboratory exercises; they take the examinations.

But various objections to student input are heard and should be recognized.

Are students able to judge good teaching or will they be overly influenced by flashy style? Will an attempt at objective evaluation turn into a popularity contest? Aren't students likely to be influenced by grades in the course? Can students accurately judge the effectiveness of teaching

while they are being taught, or does the realization of the benefits of a course only dawn on them later? Doesn't placing students in the role of evaluator inhibit trying out teaching innovations?

During the tape, our evaluation experts comment on some of these questions. Researchers have addressed most of the issues listed above, and at least tentative answers are available.

References which might be consulted include:

Kent, L. Student evaluation of teaching. In Lee, C. B. T. (Ed.), Improving college teaching. Washington: American Council on Education, 1967, pps, 312-343.

Miller, R. I. Developing programs for faculty evaluation. San Francisco: Jossey-Bass, 1974.

Murray, H. G. Evaluating University Teaching: A review of research. Toronto: Ontario Confederation of University Faculty Associations. 1980.

Nadeau, G. The evaluation of instruction: the rating questionnaire. In Knapper et al. (Eds.). If teaching is important . . . the evaluation of instruction in higher education. Toronto: Clarke-Irwin, 1977, pps. 73-128.

Questions which could be discussed among faculty members are:

(a) What are the perceived advantages and disadvantages of using student data?

(b) What empirical evidence is there to support these perceived "pluses

and minuses?"

2. The Popularity Contest

It is often felt by faculty that student ratings of instruction are based on the popularity of the instructor; that is, the degree to which the students "like" the professor, regardless of the quality of the teaching. The danger anticipated is that instructors will attempt to improve their classroom "performance," to the detriment of the content being presented.

It does appear, however, from the results of research on the validity (the degree to which an instrument measures what it is intended to measure) of student ratings, that students are able to judge teacher behavior, and to discriminate effective instructional techniques from "flashy" performance. In fact students, particularly at the higher levels of instruction, will rate the more difficult or more challenging courses higher, overall.

Some of the literature which addresses this validity question is:

Eble, K. E. The craft of teaching. San Francisco: Jossey-Bass, 1976.

Ware, J. E. and Williams, R. G. The Dr. Fox Effect: A study of lecture effectiveness and ratings of instruction. Journal of Medical Education, 1975, 50, 149-16.

Williams R. G. and Ware, J. E. An extended visit with Dr. Fox: Validity of student satisfaction with instructor ratings after repeated exposures to a lecturer. American Educational Research Journal, 1977, 14, 449-457.

Questions for discussion might include:

(a) What evidence have we seen in our teaching that students respond to quality teaching rather than popularity?

(b) Are there types of questions that could be included on a rating

form which would de-emphasize the popularity aspect of evaluations?

3. Limitations and Cautions

Since student questionnaires are the most easily quantifiable and the most efficient means of collecting evaluation information, there can be a tendency to rely on student ratings as the sole criterion of instructional effectiveness, or to assume that results are as objective as the numbers imply. As important as student reactions to instruction are, there are a number of limitations, particularly when no additional information is available.

(a) There are some aspects of instruction which students are not able to judge, such as the professor's knowledge of the subject area.

(b) The use of a rating scale does not provide in-depth information, and is subject to "response set errors," i.e. the tendency for students to respond on the "agree" side of the scale, or the "neutral" category.

(c) Student ratings have been shown to be affected by a number of variables such as class size, level of instruction, subject area, sex of the instructor, and students' reasons for taking the course.

(d) The quantification which is possible, particularly in the use of mean ratings, can mislead faculty into over-interpreting differences among these numbers (e.g. a professor who receives a mean rating of 3.42 is "better" than a colleague who receives a mean rating of 3.39).

These and any other limitations perceived to be relevant in a department or faculty should be discussed openly among all professors who will be affected by the evaluation procedures.

Some general readings are available which discuss limitations and cautions:

Dowell, D. A. and Neal, J. A. A selective review of the validity of student

ratings of teaching. Journal of Higher Education, 1982, 53, 50-62.

Renner, Richard R. Comparing professors: How student ratings contribute to a decline in quality of higher education. Phi Delta Kappan, 1981, 83, 128-130.

Sheehan, D. S. On the invalidity of student ratings for administrative personnel decisions. Journal of Higher Education, 1975, 46, 599-700.

There are likely many issues which faculty will wish to discuss under the heading of "limitations and cautions." Initial questions might include:

(a) What situations in your department or faculty should be considered in terms of limitations of the student questionnaire (e.g. a wide variety of types of courses or teaching styles)?

(b) How can questionnaires be used and interpreted to acknowledge some of the deficits of this technique?

(c) What other information can be incorporated into the evaluation procedure? (Tape No. 3 in the series expands on this question).

4. Characteristics of Good Instruction

Is there such a thing as a "good teacher" regardless of the particular course, student population, etc.?

The evidence suggests that to some extent teachers who may excell in one setting or course (or even year of teaching) may not be perceived as outstanding under other conditions. For example, the lecturer who is a fine performer before large groups of students may not be at ease and effective in leading a small seminar discussion. A major reason for collecting information over time and across a number of courses is the reduced effect of this variability.

On the other hand, it would not seem inappropriate to set minimal standards for any teacher. For example, fairness in grading or attendance

at all classes might be universally accepted as attributes expected for all members of the teaching staff.

A few global characteristics of "good teaching" are mentioned in the tape. Is there research evidence to support these (and others) as key attributes?

Numerous factor analytic and other research studies provide us with data that are unusually consistent. For example, the list of traits proposed by Issacson, McKeachie, Milholland, et al. (1964) has been essentially substantiated in many later studies. A set of attributes generally accepted by students and teacher as important would be:

1. content knowledge
2. clarity - skill in explaining the subject matter
3. rapport - friendliness, sensitivity, concern for students
4. organization - well prepared, appropriate leadership or control in class
5. workload - skill in planning appropriate amount and difficulty
6. feedback - use of criticism and praise
7. interaction - amount and kind of student participation
8. evaluation - fairness in grading; explicitness in stating requirements
9. enthusiasm about the subject matter

Some of these factors may be more important than others in students' discrimination of outstanding teachers. Thus "content expertise" may be assumed as a necessary attribute for any competent professor while warmth, a sense of humor, easy interaction with students, may more readily discriminate (in students' perceptions) the "best" from the "average."

Your faculty may wish to consult one or more of the following:

Hildebrand, M., Wilson R. C., & Dienst, E. R. Evaluating University Teaching, Berkeley: University of California, Centre for Research and Development in Higher Education, 1975.

Isaacson, R. L., McKeachie, W. J., Millholland, J. E., et al. Dimensions of student evaluations of teaching. Journal of Educational Psychology, 1964, 55, 344-351.

Wotruba, T. R. and Wright, R. How to develop a teacher rating instrument: A research approach. Journal of Higher Education, 1975, 46, 653-663.

This is the most common area of faculty concern when an evaluation is being planned. Questions to consider for discussion are:

(a) What do you consider to be the essential characteristics of effective instruction?

(b) In what situations (if any) would you consider those characteristics to be less important or even irrelevant?

5. Purpose of the Evaluation

Instructional evaluation can be conducted for a variety of reasons: individual improvement of teaching or a course, promotion or tenure decisions, other administrative decision-making such as curriculum or program revision, and the provision of information to students for course selections. Although these distinctions are not always clear in practice, it is generally the case that evaluations conducted for different purposes require different information. For example, the individual interested in improvement will be concerned with describing specific weaknesses which can be changed; the promotion and tenure committee will be interested in a comprehensive assessment over time, with general judgements of overall quality; and the student selecting a course will want to have information about workload, difficulty level, and type of examinations. As is emphasized throughout this

series of videotapes, the purpose of the evaluation should be made clear before information-collecting techniques are developed.

References which may be consulted to clarify these distinctions included:

Geis, G. L. Evaluation: definitions, problems and strategies, In Knapper, C. K. et al (Eds.) If teaching is important, Toronto, Clarke, Irwin & Co., 1977.

Glasman, N. S. and Gmelch, W. H. Purposes of evaluation of university instructors: Definitions, delineations and dimensions. Canadian Journal of Higher Education, 1976, 6, 37-55.

Questions which faculty should discuss may be:

(a) What is prompting the evaluation?

(b) Will the evaluation be primarily for improvement purposes, for administrative decision-making, or for other purposes?

(c) What types of changes or decisions are likely to be based on the evaluation results?

6. Restrictive Influences

One of the dangers of student questionnaires often perceived by faculty is that they will become a restrictive influence on teaching. That is, a professor may not wish to risk the implementation of an innovative teaching method, particularly if students are already providing positive ratings. There is, of course, the possibility of a new instructional technique receiving lower student ratings when it is first implemented: the instructor is likely to be in the process of experimentation and change during the initial stages. Related to this fear is the one that instructors will attempt to "teach to" the questionnaire; for example, to allow a great deal of discussion (if questionnaire items pertain to discussion) whether or not it is appropriate to

the situation.

This issue is best resolved by allowing individuals to choose their own questionnaire items, emphasizing those aspects of the instruction that they consider important, and taking into account innovative techniques being implemented in a course. Specific instruments for the evaluation of innovative teaching methods should be considered.

Discussion questions might include:

(a) What non-traditional instructional techniques are now being employed or being planned?

(b) What evaluation procedures could be utilized that would not restrict the present or planned innovations?

7. Standardized Versus Individualized Questionnaire

In the last fifteen years the use of student questionnaires has become so widespread that many published rating forms are available in the literature. Faculty members embarking on an evaluation plan have a choice of buying a questionnaire (or a "bank" of items) or developing their own forms.

The standardized form has the advantage, often, of having established reliability and validity and otherwise "technically good" items. On the other hand, the items may not be appropriate to a specific instructional situation, and faculty members do not have the opportunity to become involved in the development process. However the construction of individualized questionnaires involves a commitment of time and effort. Faculty members must become familiar with the basic technical techniques of item writing, and be prepared to make several revisions once a questionnaire has been developed.

In order to make the decision of standardized versus individualized

questionnaires, two areas should be reviewed: a sample of existing published forms, and guidelines for item writing.

References that may be consulted include:

Abrami, P. and Murphy, V. Catalogue of Twelve Systems for Evaluating Teaching. Montreal: McGill University, Centre for Teaching and Learning Services, Fall, 1981.

Henerson, M., Morris L. & Fitzgibbon, C. How to measure attitudes. Beverly Hills, California: Sage Publications, 1978.

McCallon, E. & McCray, E. Designing and using questionnaires. Austin, Texas: Learning Concepts, 1975.

Faculty members may wish to discuss the following questions:

(a) Are the reviewed published questionnaires appropriate to the instructional situations?

(b) How much time and effort can be committed to the development of instruments?

8. Technical Matters on Construction

The construction of a questionnaire involves at least five stages: (a) listing the aspects of instruction (e. g. course organization, lecture skill, evaluation of student learning) to be assessed; (b) selecting a rating scale, such as "strongly agree" to "strongly disagree" (c) writing clear concise statements or questions that refer to only one instructor skill or course characteristic; (d) collecting evidence that the items accurately measure what they were intended to measure, and (e) revising items that are flawed. It will be necessary, for each of these stages to refer to articles on questionnaire construction or to consult with an evaluation specialist if this resource is available.

References that should be consulted include:

Berdie, R. T. and Anderson, J. F. Questionnaires: design and use. Metuchen, New Jersey, 1974.

McCallon E. and McCray, E. Designing and using questionnaires. Austin, Texas: Learning Concepts, 1975.

Discussion should center on the following questions:

(a) What aspects of instruction will be evaluated, and will these vary across courses?

(b) What procedures can be used to evaluate the questionnaire itself?

9. Norming, Comparisons, Stratification

One of the most common questions asked when a faculty member or administrator is considering the interpretation of student questionnaire results is: can I compare these results with those of other professors? The individual instructor is concerned with "how he stands" relative to others, and the administration is often facing decisions related to the relative merit of professors or courses. This is a complex question, and is one which is at least partially dependent on the purpose of the evaluation. When the primary goal is that of individual improvement, it may be most appropriate to examine changes over time, or relative strengths and weaknesses for each individual. However, the administrator may need to consider the relative effectiveness of courses or instructors, in which case, comparisons will become an important issue. Research has shown that various course and instructor characteristics will affect the way in which students rate instruction; for example, larger classes and lower level classes tend to receive lower ratings than smaller, upper level classes. It is important then, when comparisons are made, that the appropriate comparison group be used, i.e., classes of similar size, level, and subject area.

We have suggested above (No. 6) that each professor might choose his

or her own items for a questionnaire. How then can comparisons be made across questionnaires? One frequent solution is to mandate the inclusion of the same sub-set of items on all questionnaires, usually global items that attempt to capture overall or summary opinions of the students.

Some relevant research is summarized by:

Cranton, P. A. Statistical Analyses of Evaluation Results. Paper presented at the Annual Meeting of the American Educational Research Association, New York, 1982.

Hoffman, R. G. Variables affecting university student ratings of instructor behavior. American Educational Research Journal, 1978, 15, 287-300.

Marsh, H. W. The influence of student, course, and instructor characteristics in evaluations of university teaching. American Educational Research Journal, 1980, 17, 219-238.

Questions which a department or faculty should discuss are:

(a) Are comparisons among professors necessary, and if so, why?

(b) What use would be made of comparative data?

(c) If comparisons are required how should professors and courses be grouped or stratified to allow a meaningful interpretation of the results?

10. Advantages and Disadvantages of Numerical Data

Student questionnaires most often yield numerical results, either frequencies of responses for each point on the scale of each item, average responses for each item, or even averages across groups of items or all items. At times department group means, or other "norms" are provided to allow comparisons to be made. A numerical presentation of questionnaire results is a convenient and efficient method of summarizing evaluation results. However, there are dangers involved in the use of numbers. Several sources of bias and error in student questionnaires have been discussed.

When results are presented in numerical form, there is a natural tendency to overlook possible errors and to assign too much importance to slight differences in numerical value. For example, an individual might believe that an average rating of 3.6 is better in a meaningful way than an average rating of 3.4 when this difference is actually due to one extreme student response.

Questions which faculty members should consider are:

(a) Should evaluation results be presented in numerical form alone, in a descriptive format, or in some combination of the two?

(b) How can it be assured that faculty members and administrators do not over-estimate the precision of numerical data or the value of numerical differences?

11. Administration of the Questionnaire

As with any other data collection technique, the actual method of administration of the questionnaire can influence the responses obtained. It has been shown, for example, that the directions given to students (particularly whether the questionnaire is for improvement or administrative use) will affect the ratings given. Other relevant variables may include course activities preceding and following questionnaire administration, and the degree to which confidentiality is measured.

Some further readings on this issue are:

Nadeau, G. Student evaluation of instruction: the rating questionnaire. In Knapper, C. et al (Eds.). If Teaching is Important . . . . , Toronto: Clarke Irwin, 1977, p. 91.

Seldin, P. Successful faculty evaluation programs. New York: Coventry Press, 1980, p. 61-62.

Faculty members who are designing an evaluation should consider:

(a) What method of questionnaire administration will be used?

(b) What are the possible effects of the planned administrative technique on student responses?

12. Sampling

When only one assessment is made of an individual or a situation, there is always the possibility that this assessment is not a representative sample of all the usual activities in the situation. In evaluating instruction it should be recognized that a professor may change over time or across various types of courses. It is important then, if one wishes to make a general statement about the quality of teaching or of a course, that a sample of instructional activities over time and over courses be taken.

The following question should be discussed by faculty:

(a) How can an evaluation plan be designed which will adequately sample instructional activities?

13. Students Change Over Time

Almost everyone can recall an example of a professor or a course not being appreciated at the time of teaching but later being seen as very valuable. Faculty members often argue that the effects of a course can only be evaluated long after the instructional process has terminated, perhaps when the students have embarked on relevant careers, or have advanced in their studies. This may be true, of course, for individual students; however research has shown that students in general tend to rate instruction in a similar way after the course has been completed.

The following research can be consulted:

Centra, J. A. The relationship between student and alumni ratings of teachers. Educational and Psychology Measurement, 1974, 34, 321-325.

Marsh, H. W. The validity of students' evaluations: classroom evaluations of

instructors independently nominated as best and worst teachers by graduating seniors. American Educational Research Journal, 1977, 14, 441-447.

Faculty may wish to discuss the questions:

(a) Are there particular reasons to believe that students cannot evaluate the effectiveness of a course until a later point in time?

(b) How could such an assumption be verified?

14. Improvement vs. Administrative Use of Data

The evaluation process can vary considerably depending on the use which will be made of the information collected. The content, timing, analysis and interpretation of questionnaire data is dependent on the purpose of the evaluation.

For administrative purposes, the questionnaire should contain overall judgements of the quality of the course or the instructor. The rating form is usually given near the end of the course, and analysis and interpretation often include some comparisons, either within areas on the form, or across professors or courses.

When improvement is the purpose of the evaluation, questionnaire items should address more specific aspects of the course, or teaching skills, particularly ones which are amenable to change. The questionnaire is usually administered earlier in the course in order to allow changes to be implemented. Although comparisons may be made among questionnaire items in order to identify weaknesses, it is rarely relevant to make comparisons to other courses.

In addition to the practical issue of obtaining relevant information from the questionnaire, research has shown that the stated purpose of the rating form will affect student responses.

The references given for item 5 (Purpose of the Evaluation) discuss this distinction.

Faculty members who are planning an evaluation should consider the following questions:

(a) Will the evaluation be primarily for improvement, or primarily for administrative decision-making?

(b) What affect does this stated purpose have on the content of the questionnaire, and on the information collecting process?

15. Student Course Selection

One use of evaluation information that is sometimes overlooked is student course selection. Student groups will often organize and conduct their own evaluations, publishing an "anti-calendar" or a student guide. However, it may be the case that students alone do not have sufficient resources or expertise to conduct a fair or objective assessment of the quality of the instruction. It may be worthwhile, then, for faculty to consider this third use of student questionnaire results.

As was discussed under improvement versus administrative uses, different evaluation purposes affect the nature of the questionnaire. When students are using the results for course selection they are often interested in information that is not necessarily a part of other types of evaluations. For example, emphasis may be placed on workload, difficulty level, and the nature of assignments, readings and evaluation procedures. It should be pointed out that students do not tend to give higher overall ratings to courses that they perceive as less demanding, particularly at the higher levels.

For further information, the faculty can consult the following references:

Menges, R. J. The new reporters, students rule instruction. Evaluating Learning and teaching: New Directions in Higher Education. San Francisco: Jossey-Bass, 1973.

Discussion questions may include:

(a) Should a separate evaluation for student course selection be conducted?

(b) If not, should all or part of the questionnaire results obtained for another purpose be made available to students? Or could additional items be added to a questionnaire designed for another use, and those results made "public?"

16. Over-Evaluation

Three different purposes for conducting evaluations have been discussed, with a different content and, possibly, timing for each. Students in one course could be faced with three different questionnaires over a semester, and a student enrolled in five courses might theoretically be asked to complete fifteen rating forms in three months. One might well ask whether or not this situation would have a detrimental effect on the quality of the results: at what point will students begin to take the completion of questionnaires less seriously?

Three points should be considered here. First, it is often possible to combine items for different purposes into the same form; one might have, for example, a questionnaire with two sections, one for administrative use, the results of which are only available to the instructor and the department chairman, and one section for student course selection, which is made public. Second, it is often not necessary to repeat all evaluations in each course for each professor. A fair sampling of teaching activities will often provide sufficient information for each of the described uses. Finally, when students

are clearly told the purpose of the evaluation and when they see changes being made in a course or program as a consequence of the evaluations, there will be a greater tendency to pay serious attention to the completion of rating forms.

Faculty may wish to discuss the following questions:

(a) In order to avoid over-evaluation, can questionnaire items for different purposes be presented in separate sections of the same form?

(b) Would it be possible or desirable to sample teaching activities over time and courses? If so, how could this be done in a representative way?

Summary: General Considerations for the Discussion of Student Questionnaire Use

There is a tendency to focus on the student questionnaire in the evaluation of instruction simply because it is the most efficient, best researched, and most commonly used evaluation instrument. As is emphasized in the other tapes in this series, many sources of information are available to the evaluation, and the other aspects of the evaluation process (e. g. planning, interpreting results) should be considered in detail.

Some general considerations are listed below, it is strongly recommended that they be discussed, and that the remaining tapes in this series be viewed in order to maintain a comprehensive perspective on the evaluation process.

1. The student rating form usually provides information about only one aspect of instruction: the professor's performance in the classroom. Students are generally unable to assess the instructor's knowledge of the subject area, preparation time, contribution to curriculum development, etc.

2. When the purpose of an evaluation is administrative decision-making, student questionnaires should never be used as the sole criterion for assessment. They exclude many components of university teaching and are potentially affected by several biases.

3. Although the precise nature of the effects is not clear, it has been shown that variables which are probably irrelevant to the quality of the instruction (e. g. level of the course, size of the class) affect the students' ratings.

4. Even when the effect of extraneous course characteristics is taken into account, it may be misleading to compare faculty members on the basis of student ratings. It appears that the effects interact with each other and vary across subject areas.

5. It must be accepted that the final goal of instruction is to produce an increase in students' knowledge or skills. Attempts to validate student ratings in terms of student learning have met with only moderate success. Other sources of information, such as colleague's reviews of course materials or course products may be more valuable in assessing the extent to which an instructor is achieving this important goal.

6. Unless questionnaires are designed for individual faculty members, there will be some teaching methods (particularly innovative approaches) and some teaching styles which are not adequately assessed.

7. In spite of the care that is taken in the development of an instrument and the degree to which extraneous variables are considered, there is, in any number which represents a measure of an attitude or perception, some degree of error. It is not possible, therefore, to consider small differences in these numbers as meaningful, or to make any fine discriminations on this basis.

8. It is unlikely that a questionnaire will be "perfect" as drafted. It should be tried out and revised on a sample population before it is used for evaluation purposes.

Student questionnaires can provide an efficient and useful means of collecting evaluation information. It is important, however, to consider their limitations and to supplement them with other sources of information which may better describe other aspects of the instructional situation.

Tape No. 3: Other Sources of Information

Running time: 10 1/2 minutes.

Tape 3 deals with the important topic of sources of data other than student questionnaires. Like Tape 2, it utilizes segments of interviews with evaluation specialists. Probably it should be used following Tape 2. It is important to use it sometime during discussions of evaluation in order to balance the frequent over-emphasis on student questionnaires as the primary or even sole source of information.

1. Other than responses on questionnaires what kinds of information can students contribute?

There are two kinds of student information to consider. The first is student reactions to the course and teacher. While the student questionnaire is almost universally used to obtain these comments, other techniques, some of which are mentioned on the tape, should be considered.

Open ended questions provide the opportunity for more varied, personal, and in-depth responses (e.g. What did you like best about this course? Least?) One interesting method is to ask each student to write a letter to a friend describing the course and indicating a recommendation about taking it. Letters can be typed to prevent disclosure of the author's identity. They are particularly useful for formative feedback (information to be used by the teacher for course improvement). The richness of the information is an advantage but the time consuming task of reading many such responses, the difficulty of interpretation by the reader, and the enormous variation in responses may discourage the use of this method.

Open discussions in class about the course may also be useful. Some professors have asked a student, colleague, or teaching specialist to conduct

such sessions and perhaps to summarize the results to preserve student anonymity. Some professors establish a class committee to report (say, once a week) on the course. Again, this is an appropriate method for formative evaluation. Not all teaching and learning is centered around the classroom or even the course. Many students comment that the most important part of college or university experience involves formal or informal contact with professors outside of class. Interviews with students or written comments by them can help in evaluating a professor's impact as thesis advisor, guide for independent study, model scholar, etc.

Information from student course guides, which are published on some campuses, may be consulted. Obviously the information in them is selected and edited but they do represent a documented set of student opinions (although possibly from a small group of students).

Other faculty members of chairpersons may initiate contacts with students or be approached by them. This source of information is liable to many kinds of biases and at least written records of such exchanges should be made if the inputs are to be used seriously as part of an evaluation.

There is a second kind of student data which is often discussed as an important input to evaluation: student achievement, or more broadly, evidence of the effect of the course on students.

There is some attraction in the statement: "the effectiveness of teaching can be judged by the amount of student learning." Regardless of the model we hold of ourselves as teacher, we all hope to produce some changes in students. <u>How</u> we measure those changes (i.e. our methods of evaluating students) might well be examined by colleagues or other specialists (see below). However that is not the focus here; rather it is on deciding whether and to what degree evidence of student achievement should serve

as part of evaluation of the professor.

Obviously, of the many factors affecting a student's achievement only some (perhaps very few) lie with the teacher. Moreover, the particular measures of achievement (e. g., final examinations) may not be valid or comprehensive ones. Nevertheless, it seems appropriate to include some measure of the impact of teaching in any profile of teacher effectiveness.

There have been a sizeable number of studies which examined the relationship between student ratings on questionnaires and achievement data. Surveying these studies one finds in most cases some correlation between the two measures. However it is not consistent or high enough to suggest substitution of one for the other. (For a review of these studies see Cohen, 1981).

Student achievement need not be thought of solely in terms of examination and test results. Student productions can be assessed (e. g. successful theses, theses which are published in refereed journals); questionnaires can be responded to by alumni, records on student entering graduate schools or obtaining employment can be kept and consulted; reports on field work supervision can be gathered from students and others at the field site.

2. Since colleagues judge our research contribution, shouldn't they also judge our teaching efforts?

While our research output is, in part, public (e.g. publications, research proposals), our teaching is generally private, at least with reference to the community of our peers.

Colleagues' judgments of our teaching may be influenced by such diverse things as hearsay, our prestige as scholar and researcher, interpersonal relationships among faculty members, etc. Rarely does one

colleague visit another's classroom and even when they do, research shows a lack of agreement among observers.

On the other hand, as pointed out in the tape and elsewhere in the Guide, colleagues may be the prime sources of information about course content and the quality of course documents such as outlines and examinations.

Colleagues may also be important in peer dyads or triads, the purpose of which is mutual support and assistance in improving instruction. See, for example, the peer teaching evaluation and improvement approach described by Sweeny and Grasha (1978).

3. How can colleagues assess the course materials when, the areas of expertise differ among members of the same Department?

First of all, there are likely to be many courses for which most members of the department have sufficient content expertise. In fact these are likely to be critical courses in a teaching program (e.g., the introductory course).

Secondly, outside experts might be used to review courses for which only one member of the Department is seen as having the necessary command of the content.

Thirdly, course materials from many different courses may share similar features which are not content specific. The adequacy of a course outline, even the comprehensiveness of a final examination, might be judged by colleagues who are somewhat acquainted with, but not expert in, a course's content.

Miller provides an interesting questionnaire for colleagues to use in the

evaluation of course materials (Miller, 1972).

4. How can self-assessment be used for evaluation purposes?

If a summative evaluation is being conducted for the purpose of aiding administrative decision making, self-assessment is probably of limited use. (However, interestingly, it is a part of promotion and tenure evaluation in about 1/3 of the colleges in the U.S.). In general the emphasis is upon self-report rather than self-assessment (what you did rather than how well you thought you did it).

On the other hand, self-assessment may be important for improvement or formative evaluation purposes. Methods of self-assessment range all the way from instructors comparing their own ratings with those of their students to elaborate systems of videotape recordings of the classroom with a subsequent self-critique. (For an interesting discussion of self-assessment, see Centra, J. A., 1979, pps. 47-72).

5. Is self evaluation the only information the professor supplies?

Often the evaluation data are incomplete until information is supplied by the professor of the course concerning intended goals, purposes, teaching style, methods of student evaluation etc.. Only the professor can tell us what he or she had in mind for the course and why certain methods were used. Having this information at hand may affect the interpretation of data obtained from other sources.

6. How well do the various measures of teacher quality correlate?

Given a variety of sources of information one might ask how well they correlate, i. e. how much agreement there is among them.

. Student ratings have been compared with ratings by others, for example, the professor's colleagues.

"Most studies of the relationship between student and

> colleague ratings have reported moderate to high correlations, although a few studies have reported very low correlations. However, the interpretation of nearly all of these studies is complicated by colleagues' lack of opportunity to observe teaching, nonindependence of colleagues and student raters, and unreliability of (i.e., lack of agreement among) college ratings." (Murray, p. 30).

. The relationship of student and administrator ratings appear to yield low to moderate correlations.

. Self-ratings show little correlation with student ratings, tending, as might be expected, to be considerably higher.

. Ratings by alumni (and by trained classroom observers) tend to correlate highly with student ratings. This is particularly interesting in light of the statements faculty members often make to the effect that students will only realize the high quality of teaching years after graduation.

Murray comments:

> "The most likely reason that student ratings correlate higher with alumni and external observer ratings than with colleague and administrator ratings is that students, alumni and external observers base their ratings on actual classroom observations, whereas colleagues and admnistrators presumably base their ratings on something (else)" (p. 32).

Lack of correlation may not suggest the invalidity of any source; it simply may arise from different sources attending to different aspects of the professor-as-teacher.

Despite the problems involved in judging teaching performance on the basis of student achievement (discussed above) there is some evidence of positive correlation between ratings and achievement. However, the results of more than 40 studies in this area present a picture of inconsistency. The scale seems tipped toward positive correlation but more research is necessary before one can draw conclusions with confidence.

7. In considering sources for data for evaluation, what are some options available at your institution?

On the tapes and in this Guide numerous sources of information have been proposed. It might be helpful to summarize these, some of which are examined in Tape 4.

John Centra (1977) lists these possibilities:

1. systematic student ratings
2. informal student opinions
3. colleague ratings based on classroom visits
4. colleague opinions based on evidence other than classroom visits
5. student performance on examinations
6. peer evaluation of course outlines and evaluations
7. chairman's evaluation
8. dean's evaluation
9. committee's evaluation
10. self-evaluation
11. long-term follow-up of students' performance
12. alumni opinions or ratings
13. level of student enrollment in elective courses taught by teacher being evaluated

14. analysis of videotape of classroom teaching
15. participation of faculty members in teaching improvement activities such as workshops or inservice training programs.

Each should be considered carefully in terms of the purpose of the evaluation effort and the feasibility of implementing the collection of information.

References

Centra, J. A. How universities evaluate faculty performance: a survey of department heads. GREB Research Report 75-56R. Princeton, New Jersey: Educational Testing Service, 1977.

Centra, J. A. Determining faculty effectiveness. San Francisco: Jossey-Bass, 1979.

Cohen, P. Student ratings of instruction and student achievement: A meta-analysis of multisection validity studies. Review of Educational Research, 1981, 51, 281-310.

Miller, R. I. Evaluating faculty performance. San Francisco: Jossey-Bass, 1972.

Murray, H. G. Evaluating University Teaching: A review of research. Toronto: Ontario Confederation of University Faculty Associations. 1980.

Pascal, C. E. Evaluating academic functions outside the classroom. In C. K. Knapper, G. L. Geis, C. E. Pascal and B. M. Shore (eds.) If teaching is important . . . . . the evaluation of instruction in higher education. Toronto: Clarke, Irwin & Company, 1977.

Sweeny, J. M., & Grasha, A. F. Improving teaching through faculty

development triads. Improving College and University Teaching Yearbook. Corvallis: Oregon State University Press, 1978.

Tape No. 4: Putting It All Together

Running time: 36 1/2 minutes.

Presented as a simulated interview, this tape demonstrates a suggested procedure for designing an evaluation plan. The procedure is based on the McGill Evaluation System (Cranton, 1979), however the key components are similar to these advocated by many evaluators.

Most of the basic philosophical and research issues underlying the steps in planning an evaluation have already been discussed in the tapes and the previous sections of this guide. This section will provide a brief description of the content of the steps, with some comment on matters which may be of concern to your faculty.

The first step in planning any evaluation is to clarify the purpose. As has been pointed out throughout this series, the purpose of the evaluation (improvement, administrative decision-making, student course selection) affects most aspects of the plan. In our simulation, the client is concerned with administrative decision making; this concern structures the remainder of the consultation to some extent. Had a different purpose been selected, other questions would have been raised.

The second planning step is the selection of the aspects of instruction to be evaluated. Partially dependent on the purpose of the evaluation, and partially on the details of the instructional situation, particular teaching skills or course or program components are selected (e. g. lecturing ability, textbooks, assignments, program goals) for investigation. There may be some common, agreed-upon, important characteristics within a department, or individuals may find that their courses or teaching styles are unique.

The third step in the process of planning an evaluation is to select the

appropriate sources of information and the techniques for gathering that information. These topics are dealt with in considerable detail in Tape 2 (The Student Questionnaire) and Tape 3 (Other Sources of Information). Again, it should be emphasized, particularly for administrative evaluations, that more than one source of information (e. g. students, peers) should be utilized, as a check for possible biases in any one source.

The fourth and most difficult step is the setting of criteria for decision-making or change. When are evaluation results "good"? How can cut-off points be determined? Extensive discussion among faculty members is usually required at this stage of the planning, and the assistance of a consultant can be useful. In the videotape simulation, the client returns to this issue frequently, as is usually the case in the evaluation process. Unfortunately, no one set of guidelines can be provided, as the decisions are dependent on the purpose of the evaluation, the nature of the instructional situation, and the priorities and values of the department or faculty member.

The first four steps in the evaluation procedure constitute the planning stages, and the final two the actual implementation of the evaluation. Step 5 is the collection, summary, and interpretation of the information. When the planning steps have been followed, including criteria setting, this step is relatively straightforward. As has been mentioned in the previous tapes, care must be taken in questionnaire administration (e. g., allowing anonymous responses) in order to minimize possible biases, and also in the interpretation of numerical data (e. g., avoidance of the over-emphasis on the value of numbers). Similar cautions should be exercised in the collection and interpretation of information from other sources.

The sixth and final step in the evaluation process involves making decisions or changes based on the collected results. Again this step follows

naturally from earlier planning. It is important, however, that resources are available for the implementation of change. It is not useful, for example for a faculty member to realize that a teaching skill needs improvement if there are no guidelines as to how that skill can be improved. Similarly the collection of information for administrative decision making will not be useful if there is no mechanism for its consideration in the decision-making process. Materials and resources must be made available for the evaluation itself; they should also be available for individuals who are involved in the utilization of evaluation information.

Summary

Repeatedly during the simulation on this tape the professor remarks that carrying out each step of the evaluation involves a good deal of time and effort. There is no denying that. We try to respond with some reasons why the effort is worthwhile. In your discussions it is likely the same objection will be raised. In some cases the motivation for the objection may lie in basic issues and concerns about evaluation. Tape 1 attempts to open up further discussion on these matters. Also faculty members may feel that they lack adequate technical knowledge and skill to carry out the steps of a systematic evaluation. Tapes 2 and 3 provide some background information and offer leads to further study. A specialist on campus or a consultant brought to campus can provide important assistance. Finally, faculty members may object to the amount of their time that will be required; professors typically work long days and complain that they lack time to carry out the many things they want to do. Allotting available time among various activities involves setting priorities. The faculty must agree that

systematic evaluation is a high priority activity or the effort will quickly bog down.

The time spent on planning and implementing evaluation might indeed be excessive if it resulted only in a one-shot summary ranking of professors. But we would strongly argue that the process itself illuminates. The department continues to examine its purposes and its curriculum offerings; it looks at its relationship to its students; it reviews work loads and priorities. Individual professors examine their roles as teacher; students can reflect on themselves as learners. In short: continuing attention to evaluation, incorporating the topic into the agenda of the department can be the catalyst for serious reflection, growth and renewal.

References

Centra, J. (Ed.). Determining Faculty Effectiveness. San Francisco: Jossey-Bass, 1979.

Cranton, P. A. McGill evaluation system. User Guide and Resource File Bibliography. Montreal: Centre for Teaching and Learning Services, McGill University, 1982.

Murray, H. G. Evaluating university teaching: a review of research. Toronto: Ontario, Confederation of University Faculty Associations, 1980.

Rose, C. Stalking the perfect teacher, Chronicle of Higher Education, Oct, 1981, 63(2), 128-130.

Shore, B. M. et al. Guide to the teaching dossier: its preparation and use. Ottawa: Canadian Asscciation of University Teachers 1979, Reprinted from (CAUT Bullentin) also available from Centre for teaching and Learning Services, McGill University.

A PORTABLE 5-FOOT SHELF ON EVALUATING TEACHING

Any selection of books for a basic library on evaluating teaching can be criticized for omitting some items others would include and for reflecting the biases of the selectors. Nevertheless we offer our selection here. We have limited it to a few books (or book-like publications) which we think any department should have access to when discussing and developing an evaluation system.

. Centra, J. A. Determining Faculty Effectiveness. San Francisco: Jossey-Bass, 1980.

. Cranton, P. A. McGill Evaluation System: User Guide and Resource File Bibliography. Montreal: Centre for Teaching and Learning Services, McGill University, 1982.

. Doyle, K. O. Jr. Student Evaluation of Instruction. Lexington, Mass: Heath, 1975.

. Dressel, P. L. Handbook of Academic Evaluation: Assessing instructional effectiveness, student progress and professional performance for decision-making in Higher Education. San Francisco: Jossey Bass, 1976.

. French-Lazouik, G. Evaluation of College Teaching: Guidelines for Summative and Formative Procedures. Washington, D.C.: Association of American Colleges, 1975.

. Hildebrand, M., Wilson, R. C. and Dienst, E. R. Evaluating University Teaching. Berkeley: The University of California. Centre for Research and Development in Higher Education 1975.

. Kerlinger, F. N.(Ed.). Review of Research in Education, Vol. 3, Itasca, Illinois: Peacock, 1975. See: Kulik, J. A. and McKeachie, W. J. "The evaluation of teachers in higher education."

. Knapper, C. et al (eds). If teaching is important. . . .the evaluation of instruction in higher education. Toronto: Clarke, Irwin 1979.

. Miller, R. I. Developing Programs for Faculty Evaluation: A Sourcebook for Higher Education. San Francisco: Jossey-Bass, 1974.

. Miller, R. I. Evaluating Faculty Performance. San Francisco: Jossey-Bass, 1972.

. Millman, J. (Ed.). Handbook of Teacher Evaluation, Beverly Hills: Sage Publications, 1981.

. Murray, H. G. Evaluating University Teaching: A Review of Research. Ontario: Confederation of University Faculty Associations, 1980.

. Seldin, P. Successful Faculty Evaluation Programs. New York: Coventry Press, 1980.

. Travers, R. M. W. (Ed.). Second Handbook of Research On Teaching. Chicago: Rand McNally, 1973. (See: McNeil, J. D. and Popham, W. J. "The Assessment of teacher competence," and Cohen, A. M., Trent, J. W. and Rose, C. "Evaluation of Teaching.")